

Semiparametric Estimation of Symmetric Mixture Models with Monotone and Log-Concave Densities

Xiao Pu

Ery Arias-Castro

Abstract

In this article, we revisit the problem of fitting a mixture model under the assumption that the mixture components are symmetric and log-concave. To this end, we first study the nonparametric maximum likelihood estimation (NPMLE) of a monotone log-concave probability density. By following the arguments of [Rufibach \(2006\)](#), we show that the NPMLE is uniformly consistent with respect to the supremum norm on compact subsets of the interior of the support. To fit the mixture model, we propose a semiparametric EM (SEM) algorithm, which can be adapted to other semiparametric mixture models. In our numerical experiments, we compare our algorithm to that of [Balabdaoui and Doss \(2014\)](#) and other mixture models both on simulated and real-world datasets.

1 Introduction

Let us assume that X_1, \dots, X_n are independent and identically distributed from a mixture distribution, with probability density function (pdf) $g(x)$ given by

$$g(x) = \sum_{j=1}^k \pi_j f_j(x), \quad \sum_{j=1}^k \pi_j = 1, \quad x \in \mathbb{R}, \quad (1)$$

where the number of components $k \geq 2$ is fixed and known, while the mixture proportions $\pi_j \geq 0$ and the pdf's f_j are unknown and have to be estimated. Such mixture distributions are very common in statistical modeling, and have been extensively studied ([Lindsay, 1995](#); [McLachlan and Peel, 2004](#); [Titterton et al., 1985](#)).

Note that the model (1) is not identifiable without additional constraints. To make the model identifiable, [Bordes et al. \(2006\)](#) and [Hunter et al. \(2007\)](#) propose a location-shifted semiparametric mixture model:

$$g(x) = \sum_{j=1}^k \pi_j f(x - \mu_j), \quad \sum_{j=1}^k \pi_j = 1, \quad x \in \mathbb{R}, \quad (2)$$

where $\mu_j \in \mathbb{R}$ and f is assumed symmetric (i.e., even, $f(x) = f(-x)$ for all $x \in \mathbb{R}$). These authors show that the parameters $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$ and f are uniquely identifiable when $k = 2$ (up to label-shifting) as long as $\pi_1 \neq 1/2$. Furthermore, [Hunter et al. \(2007\)](#) showed that for $k = 3$, the parameters are uniquely identifiable except when $\boldsymbol{\pi}$ and $\boldsymbol{\mu}$ take values in a particular set of Lebesgue measure zero, conjecturing that a similar result may be shown for general k . Although both [Bordes et al. \(2006\)](#) and [Hunter et al. \(2007\)](#) propose methods for estimating the parameters

Both authors are with the Department of Mathematics, University of California, San Diego, USA. Contact [Xiao \(Victor\) Pu](#) or [Ery Arias-Castro](#). This work was partially supported by a grant from the US Office of Naval Research (N00014-13-1-0257) and a grant from the US National Science Foundation (DMS 1223137).

in (2), these methods are inefficient and not easily generalizable beyond the case $k = 2$. Bordes et al. (2006) use the so-called minimum contrast method to estimate π and μ , and use a kernel density estimation (KDE) approach which involves a model selection procedure to choose the tuning parameter. Hunter et al. (2007) employ a generalized Hodges-Lehmann estimator to estimate μ and achieve a better rate of convergence. However, their estimator for f is not guaranteed to be a density. Bordes et al. (2007) propose a stochastic EM-like estimation algorithm which does not possess the monotone property of a genuine EM algorithm. Model (2) was also studied more recently by Butucea and Vandekerkhove (2014) and Balabdaoui and Doss (2014). Butucea and Vandekerkhove propose \sqrt{n} -consistent M-estimators based on a Fourier approach. Balabdaoui and Doss adopt the estimators for π and μ from Hunter et al. (2007) and then estimate the density f via maximum likelihood assuming it is log-concave. Note however that, combined, their estimators for $\{\pi, \mu, f\}$ are not obtained by maximizing the likelihood.

In this paper we consider fitting a more general mixture model where the mixture components may have different shape:

$$g(x) = \sum_{j=1}^k \pi_j f_j(x - \mu_j), \quad \sum_{j=1}^k \pi_j = 1, \quad x \in \mathbb{R}. \quad (3)$$

We assume that each f_j is symmetric and log-concave. We propose a direct maximum likelihood approach and design a genuine EM algorithm with the usual monotonicity property.

Chang and Walther (2007) have studied a similar mixture model under the assumption that each f_j is log-concave but not necessarily symmetric — obviously, the presence of the location parameter μ_j becomes redundant in that case and the model they consider is really (1) with the assumption of log-concavity. The assumption of symmetry is, however, popular, and with that assumption, for each j ,

$$f_j^+ := 2f_j \mathbb{1}_{[0, \infty)} \quad (4)$$

is a monotone log-concave density on $[0, \infty)$. Monotone densities have been used in a variety of applications and their maximum likelihood estimation was first studied by Grenander (1956). Log-concave densities have also been successfully used in nonparametric modeling and their maximum likelihood estimation has been extensively studied in the literature (Balabdaoui, 2004; Balabdaoui et al., 2009; Doss and Wellner, 2016a; Dümbgen and Rufibach, 2009; Rufibach, 2006). However the study of monotone and log-concave densities is what is required to understand the properties of our present model (3). In Section 2 we prove some basic properties for this class of densities such as uniform consistency of the MLE by simply following the existing literature, and in particular the work of Rufibach (2006).

In Section 3 we propose a genuine EM algorithm for fitting the mixture model (3). The algorithm includes a step where the monotone and log-concave MLE for f_j^+ is computed. To do so we apply the method¹ of Doss and Wellner (2016b) designed for computing the log-concave MLE with a fixed mode — the mode is of course set to 0 in our case. We note that Balabdaoui and Doss (2014) use the same routine in the numerical implementation of their method.

In Section 4 we apply our model to clustering problems and compare our approach with that of (Chang and Walther, 2007) (without symmetry) and that of (Balabdaoui and Doss, 2014) (without a monotone EM), as well as a Gaussian mixture model (GMM), on both synthetic and real-world datasets, in terms of misclassification errors, posterior errors, and achieved likelihood.

¹The method is based on an active set implementation and has been implemented in the R package **logcondens.mode**.

2 NPMLE of a monotone log-concave density

This section is concerned with the estimation of a monotone log-concave density f via maximum likelihood from a given ordered sample $x_1 < x_2 < \dots < x_n$. We let F denote the distribution function corresponding to the density f and define

$$\psi(x) = \log f(x). \quad (5)$$

Requiring that f be monotone and log-concave is equivalent to requiring that ψ is monotone (non-increasing) and concave.

Based on the sample, the negative log-likelihood at f is given by

$$-\sum_{i=1}^n \log f(x_i) = -n \sum_{i=1}^n \psi(x_i). \quad (6)$$

In order to relax the constraint of f being a probability density we follow the trick used by [Rufibach \(2006\)](#) and add a Lagrange term to (6), leading to the functional

$$\Lambda_n(\psi) = -\sum_{i=1}^n \psi(x_i) + n \int \exp \psi(x) dx. \quad (7)$$

The NPMLE of f is $\hat{f}_n = \exp \hat{\psi}_n$, where $\hat{\psi}_n$ is the minimizer of Λ over class of functions on $[0, \infty)$ that are non-increasing and concave, that is

$$\hat{\psi}_n := \arg \min_{\psi \in \mathcal{MC}} \Lambda_n(\psi), \quad (8)$$

where²

$$\mathcal{MC} := \{\psi : [0, \infty) \rightarrow [-\infty, \infty) \mid \psi \text{ is non-increasing, concave, proper, and closed}\}. \quad (9)$$

The theory below results from a straightforward adaptation of the thesis work of ([Rufibach, 2006](#)) on the maximum likelihood of a log-concave density, without the additional constraint of monotonicity, published in the form of a research article in ([Dümbgen and Rufibach, 2009](#)). We do not provide proofs but rather refer the reader to that work.

The following results from an adaptation of Theorem 2.1 in ([Dümbgen and Rufibach, 2009](#)).

Theorem 1 (Existence, uniqueness, and shape). *The NPMLE $\hat{\psi}_n$ exists and is unique. It is linear between sample points and continuous on $[0, x_n]$, with $\hat{\psi}_n(x) = \hat{\psi}_n(x_1)$ for $x \in [0, x_1]$ and $\hat{\psi}_n(x) = -\infty$ for $x > x_n$.*

The following results from an adaptation of Theorem 2.2 in ([Dümbgen and Rufibach, 2009](#)).

Theorem 2 (Characterization). *Let ψ be a non-increasing and concave function such that $\{x : \psi(x) > -\infty\} = [0, x_n]$. Then, $\psi = \hat{\psi}_n$ if and only if*

$$\frac{1}{n} \sum_{i=1}^n \Delta(x_i) \leq \int \Delta(x) \exp \psi(x) dx \quad (10)$$

for any $\Delta : [0, \infty) \rightarrow \mathbb{R}$ such that $\psi + \lambda \Delta$ is non-increasing and concave for some $\lambda > 0$.

²Following the definition in [Rockafellar \(2015\)](#), a concave function f is said to be proper if $f(x) > -\infty$ for at least one x and $f(x) < +\infty$ for every x . A closed function is a function that maps closed sets to closed sets.

For $I \subset \mathbb{R}$ an interval, $\beta \in [1, 2]$, and $L > 0$, let $\mathcal{H}^{\beta, L}(I)$ be the Hölder class of real-valued functions g on I satisfying $|g(y) - g(x)| < L|y - x|$ if $\beta = 1$ and $|g'(y) - g'(x)| \leq L|y - x|^{\beta-1}$ if $\beta \in (1, 2]$, for all $x, y \in I$. The following results from an adaptation of Theorem 4.1 in (Dümbgen and Rufibach, 2009).

Theorem 3 (Uniform consistency). *Assume that $f \in \mathcal{H}^{\beta, L}(I)$ for some exponent $\beta \in [1, 2]$, some constant $L > 0$, and a compact interval $I \subset \{f > 0\}$. Then,*

$$\max_{t \in I} |\hat{f}_n(t) - f(t)| = O_{\mathbb{P}}(\log n/n)^{\beta/(2\beta+1)}. \quad (11)$$

As pointed out by Dümbgen and Rufibach (2009), this is the minimax rate for densities in that smoothness class, as shown by Khas'minskii (1979), so that, when the density is log-concave and Hölder- β (with $\beta \in [1, 2]$) in some interval, the log-concave MLE adapts to the proper smoothness in that interval. We believe the same holds under the additional constraint of monotonicity.

3 A semiparametric EM algorithm

We now consider fitting the semiparametric mixture model (3). Recalling (4), this amounts to estimating $\phi := (\boldsymbol{\mu}; \boldsymbol{\pi}; \mathbf{f}^+)$ where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k) \in \mathbb{R}^k$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ is an element of the simplex in \mathbb{R}^k , and $\mathbf{f}^+ = (f_1^+, \dots, f_k^+)$ with each f_j^+ being a monotone log-concave density on $[0, \infty)$. Under ϕ , the density of the mixture model is given by

$$g_{\phi}(x) = \frac{1}{2} \sum_{j=1}^k \pi_j f_j^+(|x - \mu_j|). \quad (12)$$

The log-likelihood associated of the sample $\mathbf{x} = (x_1, \dots, x_n)$ under parameter ϕ is thus given by

$$L(\phi) = \sum_{i=1}^n \log g_{\phi}(x_i). \quad (13)$$

It is well-known that directly maximizing $L(\phi)$ is difficult. We design an EM-type algorithm. Let $z_i = j$ when x_i was sampled from the j th component, and define

$$w_{ij} = \mathbb{P}_{\phi}(z_i = j | x_i) = \frac{\pi_j f_j^+(|x_i - \mu_j|)}{\sum_{l=1}^k \pi_l f_l^+(|x_i - \mu_l|)}. \quad (14)$$

With these particular weights, clearly,

$$L(\phi) = \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log(\pi_j f_j^+(|x_i - \mu_j|)) - C(\mathbf{w}), \quad (15)$$

where $C(\mathbf{w}) := \sum_{i=1}^n \sum_{j=1}^k w_{ij} \log w_{ij} + n \log 2$. For a set of parameters $\phi = (\boldsymbol{\mu}; \boldsymbol{\pi}; \mathbf{f}^+)$ and weights $\mathbf{w}_{*} = (w_{ij*})$, define

$$Q(\phi, \mathbf{w}_{*}) = \sum_{i=1}^n \sum_{j=1}^k w_{ij*} \log(\pi_j f_j^+(|x_i - \mu_j|)). \quad (16)$$

In an iterative implementation, assuming that $\phi_{(t)}$ denotes the set of parameters at iteration t and $\mathbf{w}_{(t)}$ the weights computed according to (14), a typical EM approach requires the maximization of $Q(\phi, \mathbf{w}_{(t)})$ with respect to ϕ . We propose alternative optimization procedure to do so.

The semiparametric EM (SEM) algorithm that we deploy is described below.

- **E-step:** Given $\phi_{(t)}$, we calculate

$$w_{ij(t)} = \mathbb{P}(z_i = j | x_i, \phi_{(t)}) = \frac{\pi_{j(t)} f_{j(t)}^+ (|x_i - \mu_{l(t)}|)}{\sum_{l=1}^k \pi_{l(t)} f_{l(t)}^+ (|x_i - \mu_{l(t)}|)}, \quad (17)$$

for $i = 1, \dots, n$ and $j = 1, \dots, k$.

- **M-step:**

- Update π

$$\pi_{j(t+1)} = \frac{1}{n} \sum_{i=1}^n w_{ij(t)}, \quad j = 1, \dots, k; \quad (18)$$

- Update μ

$$\mu_{j(t+1)} := \arg \max_{\mu} \sum_{i=1}^n w_{ij(t)} \log f_{j(t)}^+ (|x_i - \mu|), \quad j = 1, \dots, k; \quad (19)$$

- Update f^+

$$f_{j(t+1)}^+ := \arg \max_{f^+} \sum_{i=1}^n w_{ij(t)} \log f^+ (|x_i - \mu_{j(t+1)}|), \quad j = 1, \dots, k. \quad (20)$$

Since $\log f_{j(t)}^+$ is concave, the objective function in (19) is a concave function of μ , therefore Golden Section Search can be applied to solve this optimization problem. In (20), the optimization is over f^+ being a monotone and log-concave density on $[0, \infty)$. The solution corresponds to the weighted NPMLE based on data $(|x_1 - \mu_{j(t+1)}|, \dots, |x_n - \mu_{j(t+1)}|)$ and weights $(w_{1j(t)}, \dots, w_{nj(t)})$.

Remark 1. Our implementation is based on applying the function `activeSetLogCon.mode` in the R package `logcondens.mode` with mode chosen to be 0.

- **Initialization:** We initialize $\mathbf{w}_{(0)}$ and $\mathbf{f}_{(0)}^+$ at the values given by a fit of a GMM, and start with M-step first.

Our SEM algorithm has the desirable monotonicity property of a true EM algorithm (Dempster et al., 1977; Wu, 1983).

Proposition 1 (Monotonicity property). *With the same notation, $L(\phi_{(t)}) \leq L(\phi_{(t+1)})$ for all $t \geq 0$.*

Proof. In the algorithm, armed with $\phi_{(t)}$, we compute the weights $\mathbf{w}_{(t)}$ in the E-step and in the M-step we obtain $\phi_{(t+1)}$ by maximizing $Q(\phi, \mathbf{w}_{(t)})$ over ϕ . (We do the latter sequentially, first over π , then over μ , and finally over f^+ .) In particular,

$$Q(\phi_{(t+1)}, \mathbf{w}_{(t)}) \geq Q(\phi_{(t)}, \mathbf{w}_{(t)}). \quad (21)$$

The key, then, is Jensen's inequality, which implies that for a set of parameters $\phi = (\mu; \pi; f^+)$ and non-negative weights $\mathbf{w}_* = (w_{ij*})$ such that $\sum_j w_{ij*} = 1$ for all i ,

$$L(\phi) = \sum_{i=1}^n \log \left(\frac{1}{2} \sum_{j=1}^k \pi_j f_j^+ (|x_i - \mu_j|) \right) \quad (22)$$

$$\begin{aligned} &= \sum_{i=1}^n \log \left(\sum_{j=1}^k w_{ij*} \frac{\pi_j f_j^+ (|x_i - \mu_j|)}{w_{ij*}} \right) - n \log 2 \\ &\geq \sum_{i=1}^n \sum_{j=1}^k w_{ij*} \log \left(\frac{\pi_j f_j^+ (|x_i - \mu_j|)}{w_{ij*}} \right) - n \log 2 \\ &= Q(\phi, \mathbf{w}_*) - C(\mathbf{w}_*), \end{aligned} \quad (23)$$

with equality if the weights \mathbf{w}_* are the weights associated with ϕ as specified in (14). In particular,

$$L(\phi_{(t+1)}) \geq Q(\phi_{(t+1)}, \mathbf{w}_{(t)}) - C(\mathbf{w}_{(t)}), \quad (24)$$

while

$$L(\phi_{(t)}) = Q(\phi_{(t)}, \mathbf{w}_{(t)}) - C(\mathbf{w}_{(t)}). \quad (25)$$

We thus have

$$\begin{aligned} L(\phi_{(t+1)}) &\geq Q(\phi_{(t+1)}, \mathbf{w}_{(t)}) - C(\mathbf{w}_{(t)}) \\ &\geq Q(\phi_{(t)}, \mathbf{w}_{(t)}) - C(\mathbf{w}_{(t)}) = L(\phi_{(t)}). \end{aligned} \quad \square$$

4 Numerical experiments

We now consider the problem of one-dimensional clustering. We assume that the data can be clustered into k groups, fit the k -component mixture (3) as described in Section 3 obtaining $\hat{\phi}$, and assign a label to an observation x_i according to Bayes optimal rule

$$\arg \max_j \mathbb{P}(z_i = j | x_i, \hat{\phi}) = \arg \max_j \frac{\hat{\pi}_j \hat{f}_j^+(|x_i - \hat{\mu}_j|)}{\sum_{l=1}^k \hat{\pi}_l \hat{f}_l^+(|x_i - \hat{\mu}_l|)}. \quad (26)$$

We apply our SEM algorithm both on simulated and real data. In Section 4.1, we choose to simulate data from the Gaussian and Laplace mixture models used in (Balabdaoui and Doss, 2014), and in Section 4.2, we apply the SEM algorithm to the well-known Old Faithful geyser data also investigated in (Balabdaoui and Doss, 2014), and to the rainfall data studied in (Bordes et al., 2006).

4.1 Synthetic datasets

As a first example, we use a two-component Gaussian mixture to empirically check the convergence of our SEM algorithm. We first sample $n = 100$ (Figure 1) and then sample $n = 300$ (Figure 2) observations from the Gaussian mixture $0.15 \mathcal{N}(-1, 1) + 0.85 \mathcal{N}(2, 1)$ and apply the SEM algorithm to these two datasets respectively. This seems to be the most difficult situation considered in (Bordes et al., 2006). Panels (a), (b), (c), and (d) of Figure 1 show that SEM stabilizes after about 8 iterations for the three Euclidean parameters and the observed data likelihood. As expected, the achieved maximum data likelihood is monotonically increasing as a function of the number of iterations. Panels (e) and (f) show the final NPMLE for f_1, f_2, g and compare that with the truth. The NPMLE for the symmetric log-concave densities are piecewise exponential, which is consistent with what is described in Theorem 1. Figure 2 is provided to show on the improvement resulting from a larger sample size. With initialization manually set the same with that in Figure 1, the effect is visible on the recovering of f_1^+, f_2^+ and g .

We then conduct a Monte Carlo study to compare the performance of our algorithm (SEM) in clustering with the methods proposed in (Chang and Walther, 2007) and (Balabdaoui and Doss, 2014). Chang and Walther fit the simple mixture model (1) and only assume that the components are log-concave densities. Balabdaoui and Doss fit the semiparametric model (2) and employ the parameter estimators from (Hunter et al., 2007), and then assumes that both components have the same density after centering and fits that density using the symmetric log-concave density estimator. We denote these two methods by LCM and SLC respectively. We compare SEM, LCM and SLC to GMM, which serves as benchmark when the underlying model is indeed in that class.

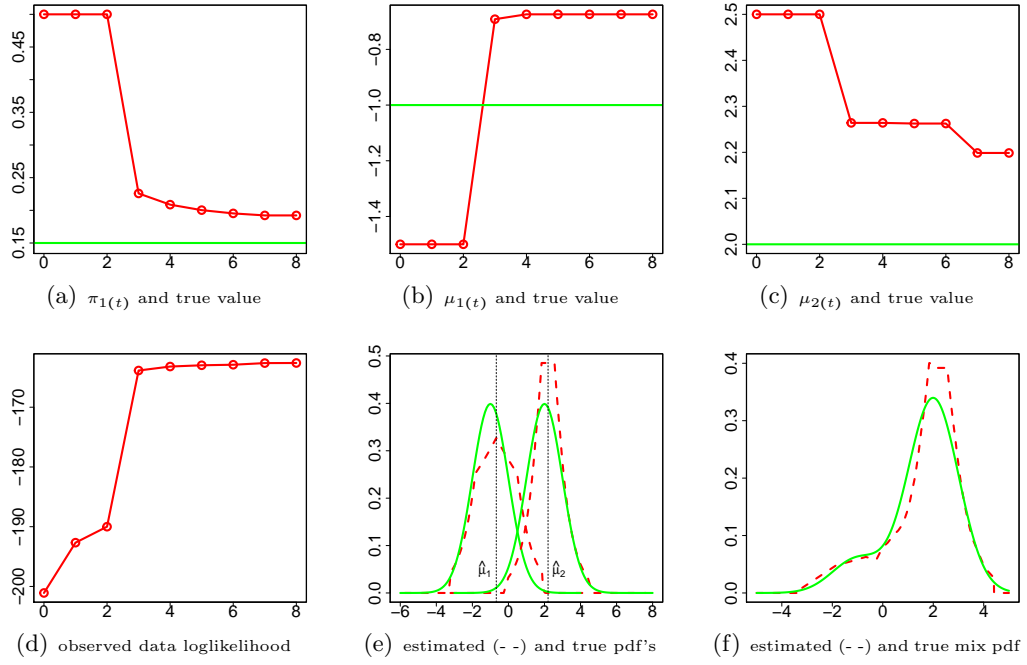


Figure 1: SEM for the Gaussian mixture with $n = 100$, $\pi_1 = 0.15$, $\mu_1 = -1$ and $\mu_2 = 2$.

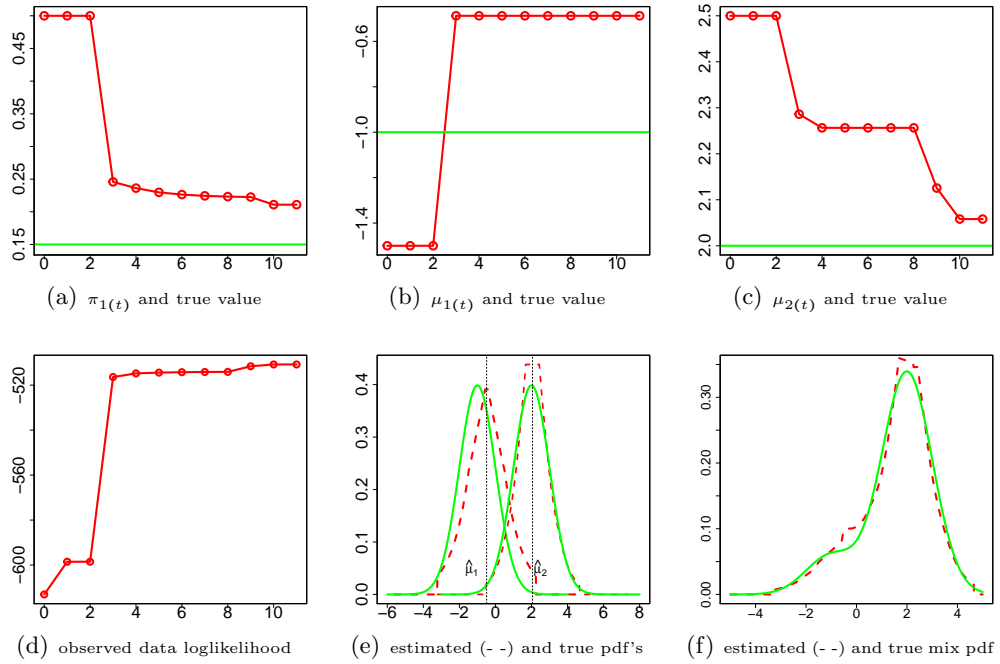


Figure 2: SEM for the Gaussian mixture, $n = 300$, $\pi_1 = 0.15$, $\mu_1 = -1$ and $\mu_2 = 2$.

	Metric	GMM	LCM	SLC	SEM
Model 1	log-likelihood	-743.2 (0.50)	-739.4 (0.51)	-1104.9 (2.06)	-738.6 (0.50)
	mis-class	122.7 (1.31)	102.6 (1.49)	174.2 (1.22)	123.7 (1.30)
	post-error	0.199 (0.003)	0.206 (0.002)	0.317 (0.001)	0.202 (0.003)
Model 2	log-likelihood	-1049.7 (0.48)	-1046.3 (0.49)	-1383.0 (2.43)	-1044.7 (0.48)
	mis-class	148.1 (1.55)	125.4 (1.91)	118.4 (0.70)	150.0 (1.54)
	post-error	0.211 (0.004)	0.255 (0.003)	0.216 (0.004)	0.283 (0.001)
Model 3	log-likelihood	-876.0 (0.67)	-869.3 (0.69)	-1293.6 (3.94)	-870.6 (0.66)
	mis-class	162.8 (1.12)	111.4 (1.23)	153.2 (1.09)	159.1 (1.18)
	post-error	0.236 (0.002)	0.244 (0.002)	0.324 (0.001)	0.234 (0.002)

Table 1: Model 1: $0.2\mathcal{N}(0, 1) + 0.8\mathcal{N}(1, 1)$; Model 2: $0.2\mathcal{N}(0, 1) + 0.8\mathcal{N}(2, 2)$; Model 3: $0.2\mathcal{L}(0, 1) + 0.8\mathcal{L}(1, 1)$. Comparison of the three different clustering methods in terms of achieved log-likelihood, number of misclassification errors, and posterior errors. The reported numbers are the average of the metrics over $R = 1000$ replications under each of the three symmetric and log-concave mixture models. Each time the sample size is $n = 500$. The numbers in parentheses are the corresponding standard errors.

We compare these methods on two Gaussian mixture models and one Laplace mixture model. The sample size is $n = 500$. Each setting is repeated 1000 times. We examine the quality of the resulting clustering in terms of the achieved data log-likelihood, the misclassification errors, and the average absolute posterior probability error used by (Chang and Walther, 2007) — all averaged over the 1000 repeats. The latter metric investigates how well a mixture clustering algorithm estimates the uncertainty for the membership assignment of each observation on population level. This metric is defined as

$$\text{posterior error} := \frac{1}{n} \sum_{i=1}^n |\hat{w}_{i1} - w_{i1}| \quad (27)$$

where \hat{w}_{i1} and w_{i1} are computed by (14) with estimators and true parameters respectively. We report the comparison results in Table 1. As can be seen from this table, GMM, LCM and our SEM algorithm clearly outperforms SLC in terms of log-likelihood and posterior error. When the mixture densities are normal, SEM performs as well as GMM, arguably the gold standard in such a situation; when the densities are Laplace, SEM slightly improves the clustering initialized by GMM. Moreover, SEM achieves significantly a higher log-likelihood for all of these 4 models when the mixture densities are normal. We also notice that SLC sometimes gives better results in terms of misclassification error, even though the posterior-error is worse.

4.2 Real datasets

In this section, we apply our new estimation approach to two different real-world datasets. Both of these datasets are included in the standard R distribution.

The first dataset consists of times, in minutes, between eruptions of the Old Faithful geyser in Yellowstone National Park. Figure 3 plots the iterations of our SEM algorithm, which is seen to converge rather quickly, in less than 14 iterations. Table 2 shows that our estimates are close to those obtained by GMM, Hunter et al. (2007), and Bordes et al. (2007), while the estimates from Balabdaoui and Doss (2014) are a bit farther away.

The second dataset is the average amount of precipitation (rainfall) in inches for each of 70 cities in the United States and Puerto Rico (McNeil, 1977). Figure 4 plots the iterations of our SEM algorithm, which is again seen to converge quickly.

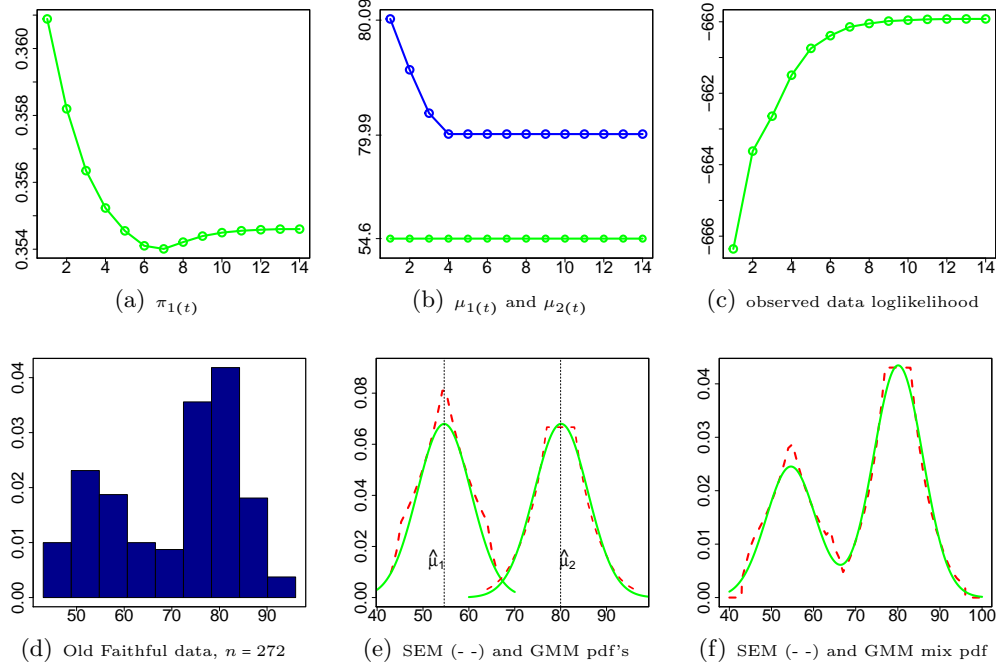


Figure 3: SEM applied to the Old Faithful waiting data.

parameters	GMM	SP	SP-EM	SLC	SEM
π_1	0.361	0.352	0.359	0.33	0.355
μ_1	54.61	54.0	54.59	55.5	54.61
μ_2	80.09	80.0	80.05	80.5	80.5

Table 2: Parameter estimates for the Old Faithful geyser waiting data, using GMM, the semiparametric estimation from Parameter estimates for the Old Faithful geyser waiting data, using GMM, the semiparametric estimation from [Hunter et al. \(2007\)](#)(SP), the stochastic EM algorithm by [Bordes et al. \(2007\)](#) (SP-EM), the symmetric log-concave mixture model by [Balabdaoui and Doss \(2014\)](#) (SLC) and our SEM algorithm.

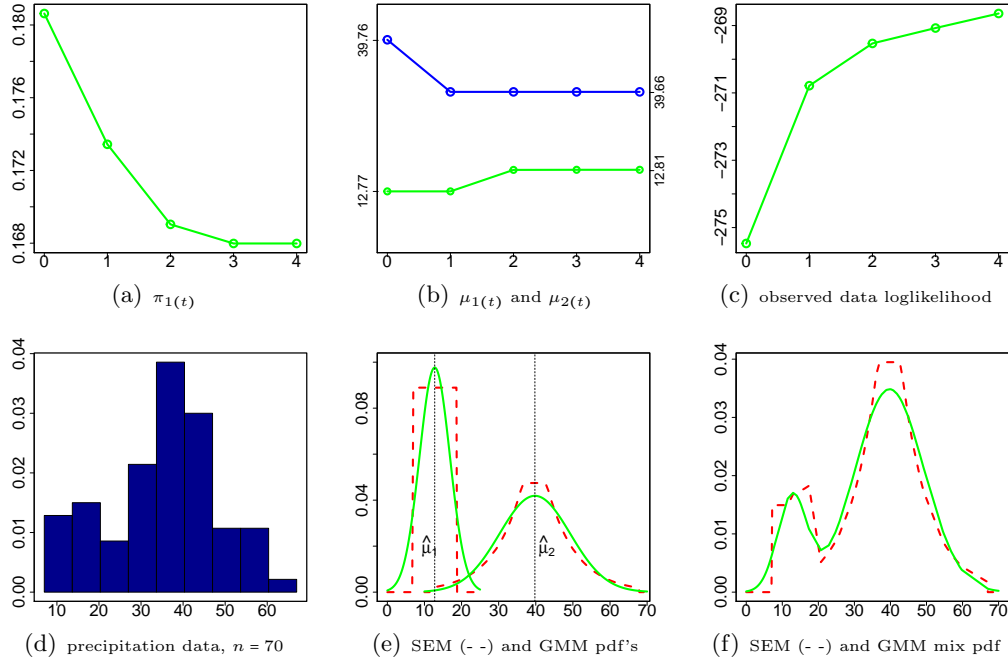


Figure 4: SEM applied to the annual precipitation data.

References

- Balabdaoui, F. (2004). *Nonparametric estimation of a k -monotone density: A new asymptotic distribution theory*. Ph. D. thesis, University of Washington.
- Balabdaoui, F. and C. R. Doss (2014). Inference for a mixture of symmetric distributions under log-concavity. *arXiv preprint arXiv:1411.4708*.
- Balabdaoui, F., K. Rufibach, and J. A. Wellner (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of statistics* 37(3), 1299.
- Bordes, L., D. Chauveau, and P. Vandekerkhove (2007). A stochastic em algorithm for a semiparametric mixture model. *Computational Statistics & Data Analysis* 51(11), 5429–5443.
- Bordes, L., S. Mottelet, and P. Vandekerkhove (2006). Semiparametric estimation of a two-component mixture model. *The Annals of Statistics* 34(3), 1204–1232.
- Butucea, C. and P. Vandekerkhove (2014). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics* 41(1), 227–239.
- Chang, G. T. and G. Walther (2007). Clustering with mixtures of log-concave distributions. *Computational Statistics & Data Analysis* 51(12), 6242–6251.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.
- Doss, C. R. and J. A. Wellner (2016a). Global rates of convergence of the mles of log-concave and s -concave densities. *The Annals of Statistics* 44(3), 954–981.
- Doss, C. R. and J. A. Wellner (2016b). Mode-constrained estimation of a log-concave density. *arXiv preprint arXiv:1611.10335*.
- Dümbgen, L. and K. Rufibach (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* 15(1), 40–68.
- Grenander, U. (1956). On the theory of mortality measurement: part ii. *Scandinavian Actuarial*

- Journal* 1956(2), 125–153.
- Hunter, D. R., S. Wang, and T. P. Hettmansperger (2007). Inference for mixtures of symmetric distributions. *The Annals of Statistics*, 224–251.
- Khas'minskii, R. (1979). A lower bound on the risks of non-parametric estimates of densities in the uniform metric. *Theory of Probability & Its Applications* 23(4), 794–798.
- Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics*, pp. i–163. JSTOR.
- McLachlan, G. and D. Peel (2004). *Finite mixture models*. John Wiley & Sons.
- McNeil, D. R. (1977). Interactive data analysis.
- Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.
- Rufibach, K. (2006). *Log-concave density estimation and bump hunting for IID observations*. Ph. D. thesis.
- Titterton, D., A. Smith, and U. Makov (1985). *Statistical analysis of finite mixture distributions*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.
- Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, 95–103.